УДК 004.7:004.8

КЛАСИФІКАЦІЯ ВУЗЬКОЦІЛЬОВИХ ТА БАГАТОЦІЛЬОВИХ ПРИСТРОЇВ ІНТЕРНЕТУ РЕЧЕЙ З ВИКОРИСТАННЯМ МАШИННОГО НАВЧАННЯ

В. В. Мельник[1,3,а], П. О. Галета[2,3,б]

[1] *Національний технічний університет України «Київський політехнічний інститут» імені Ігоря Сікорського, Фізико-технічний інститут*

[2] *Національний технічний університет України «Київський політехнічний інститут» імені Ігоря Сікорського*

[3] *Samsung R&D Institute Ukraine (SRK)*

Анотація

Через недоліки в проектуванні, реалізації та неналежному налаштуванні в мережі значно погіршується працездатність пристроїв Інтернету речей. Як наслідок, Інтернет речі стають вразливими місцями в мережі та легкими цілями для компроментації, а також можуть бути долученими до ботнет мережі. Таким чином, мережа стає вразливою до таких атак, як різноманітні DoS та DDoS атаки, сканування мережі тощо. В свою чергу, механізм класифікації пристроїв в мережі дозволяє більш гнучко будувати правила безпеки та правила QoS для локальної мережі або створювати окремі віртуальні мережі. Існуючі рішення не дають належного відсотку коректного класифікування пристроїв на вузькоцільові та багатоцільові Інтернет речі та продуктивні лише на конкретній змодельованій системі. Запропонований підхід демонструє удосконалений аналіз потоку пакетів в мережі. Отриманий результат показує працездатність методу класифікації на вузькоцільові та багатоцільові Інтернет речі на різних модельованих системах.

Ключові слова: Інтернет речі, машинне навчання, локальна мережа, інтернет-трафік, DNS

Вступ

Інтернет речі на сьогоднішній день набирають стрімку популярність через те, що ці пристрої можуть виконувати повсякденні задачі автономно, без прямого втручання людини. Згідно з прогнозами вчених [1], кількість Інтернет речей до 2020 року зросте близько до 50 мільярдів, що дає підстави говорити про початок ери Інтернет речей.

Однак з ростом популярності Інтернет речей росте і небезпека їх використання. Через масовість використання, оператори розумних середовищ стикаються з великою проблемою розпізнавання підключених пристроїв і, відповідно, із з'ясуванням коректності їх роботи. Так, наприклад, в 2017 році за допомогою торгових автоматів, що знаходилися на території університету, таку атаку зазнав університетський кампус [2]. Як результат, було пошкоджено 5000 пристроїв Інтернету речей. Тому, відповідно до звіту компанії Cisco [3], виявлення та класифікація кожного пристрою – одна із цілей для впевненості у тому, що кожен пристрій знаходиться в безпечному для нього сегменті мережі та забезпечує необхідну якість обслуговування.

Головною метою даної статті є створення ефективного механізму класифікації пристроїв в мережі за допомогою дослідження мережевого потоку пакетів від кожного пристрою (трафіку). На основі отриманих даних будується загальна картина поведінки кожного пристрою, за допомогою якої класифікатор може віднести цей пристрій до вузькоцільових або багатоцільових Інтернет речей. Даний метод може бути застосований до різних мереж, що побудовані за допомогою різних представників серед Інтернету речей.

1. Огляд існуючих рішень

На сьогоднішній день існують численні методи класифікації пристроїв інтернету речей, більшість з яких базується на характеристиках потоку пакетів.

Так, в [4] було запропоновано метод класифікації пристроїв за допомогою особливостей спектру отриманого шляхом перетворення Фур'є дискретного сигналу передачі службових протоколів, таких як ARP, mDNS, NTP. Особливістю даного методу є швидка класифікація пристроїв, що, за даними статті, складає 90 хвилин після початку спостереження. Однак недоліком є непрацездатність даного методу в мережах, відмінних від запропонованої в силу різних можливостей налаштувань кожного пристрою.

У [5] було запропоновано використовувати імена доменів, до яких пристрої звертаються в процесі комунікації. Фільтруючи надалі імена доменів, що належать виробнику пристрою, досягається досить

[а] v.melnik@samsung.com
[б] p.haleta@samsung.com

висока точність класифікації. Однак головним недоліком цього методу є те, що не всі пристрої мають зв'язок з доменами, ключові імена яких вміщують назву виробника пристрою. Ще однією особливістю є те, що вузькоцільовий і багатоцільовий пристрої можуть належати одному виробнику, таким чином використовуючи спільне ім'я домену.

У [6] було запропоновано метод класифікації за допомогою аналізу статистичних характеристик потоку пакетів кожного пристрою. Як було встановлено на практиці, даний метод є працездатним, однак в межах однієї практично змодельованої мережі.

Дана стаття зосереджена на удосконаленні останнього запрпонованого методу [6] шляхом застосування додаткових припущень, що однозначно здатні серед усіх пристроїв локальної мережі відокремити вузькоцільові та багатоцільові Інтернет речі.

2. Класифікація пристроїв Інтернету речей за допомогою поведінкових профілів

2.1. Визначення основних понять

Запропонований метод застосовується до будь-якої локальної мережі, в якій всі пристрої мають підключення до мережі Інтернет. Пристрій, що має здатність до мережевого підключення та має вбудований будь-якого роду передавач (sensor), можна віднести до Інтернету речей. Виходячи з цього, можна виокремити наступні типи пристроїв:

- Вузькоцільові пристрої: до даної групи пристроїв відносяться ресурсо-обмежені пристрої, такі як передавачі, камери з обмеженою та/або нескладною функціональністю тощо. Також дані пристрої не потребують людського втручання.
- Багатоцільові пристрої: до даної групи належать високотехнологічні пристрої з кращими апаратними ресурсами, такі як смартфони, персональний комп'ютер тощо.

2.2. Основні припущення

Припущення 1. DNS – один з найбільш популярних протоколів серед Інтернету речей. Шляхом аналізу потоків пакетів було встановлено, що кількість DNS запитів, за тривалий період спостереження, багатоцільових Інтернет речей значно перевищує кількість DNS запитів вузькоцільових Інтернет речей, причому більшість DNS запитів вузькоцільових Інтернет речей містять імена своїв виробників. Для багатоцільових пристроїв, як телефон, дана картина не є типовою. Для порівняння, результати аналізу відображені на рис. 1. Таким чином, кількість унікальних DNS запитів та кількість унікальних імен доменів, з якими встановлюють з'єднання Інтернет речі, є важливими показниками, що здатні класифікувати вузькоцільові та багатоцільові інтернет речі.

Припущення 2. Сила зв'язності багатоцільових пристроїв перевищує силу зв'язності вузькоцільових через той факт, що багатоцільовий пристрій

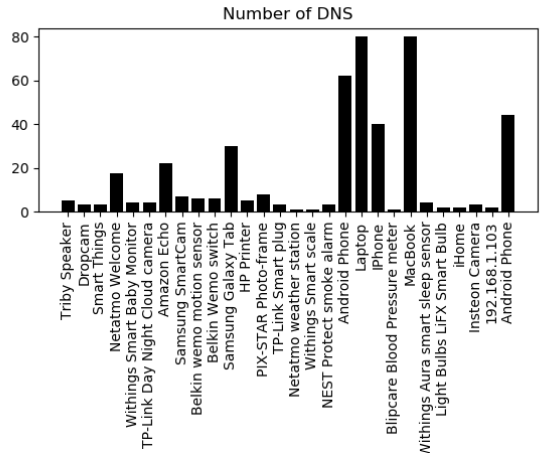

Рис. 1. Кількість DNS запитів вузькоцільових та багатоцільових пристроїв (за 1 тиждень)

здатен встановлювати з'єднання з багатьма вузькоцільовими пристроями. Сила зв'язності – індекс, що вказує, як пов'язані пристрої між собою в мережі. Як приклад: за допомогою мобільного телефона є можливість контролювати більшість сенсорів, камер, вузькоцільових акустичних систем тощо.

В даній статті для обчислення сили зв'язності кожного пристрою використовується технологія Google's PageRank [7], що аналізує орієнтований граф зв'язності в мережі, кожне ребро якого має вагу, що дорівнює кількості пакетів згенерованих кожним пристроєм. Отриманий результат (таб. 1) аналізу сили зв'язності доводить, що сила зв'язності багатоцільових пристроїв перевищує силу зв'язності вузькоцільових.

Припущення 3. Поле User-Agent, що передається в HTTP заголовку, притаманно в більшості випадків для багатоцільових пристроїв. Таким чином, проаналізувавши даний показник, можна встановити тип пристрою та зробити необхідний висновок. Дана характеристика є категоріальною та сприяє збільшенню точності у визначенні класу пристрою, але може бути використаною тільки у випадку з незашифрованим трафіком.

2.3. Опис алгоритму

Запропонований метод класифікації зображений на рис. 2. Він побудований на основі поведінкового профілю, що складається з статистичних характеристик потоку пакетів.

Типовими характеристиками потоку пакетів в даній моделі є:

1) Кількість DNS запитів.
2) Кількість невідомих зв'язків (кількість зв'язків, для яких не використовувались доменні імена).
3) Кількість типів протоколів, що були використані для комунікації.
4) Характеристики TCP з'єднань, що включають середнє значення об'єму пакетів, переданих за одну сесію та середній час однієї сесії.
5) Тип пристрою за полем user-agent.

Табл. 1. Сила зв'язності пристроїв в локальній мережі

Пристрій	PageRank
Router	0.3
Securify Almond	0.050131
Philips HUE Hub	0.041608
Samsung SmartThings Hub	0.030454
Android Tablet	0.026212
Apple HomePod	0.024202
Google Home	0.0221
Google Home mini	0.0221
Samsung SmartTV	0.020897
Sonos	0.018622
MiCasaVerde VeraLite	0.016152
Roku 4	0.016152
Roku TV	0.016152
iPad	0.016091
Wink 2 Hub	0.014857
Amazon Fire TV	0.014113
Apple TV	0.013064
D-Link DCS-5000L Camera	0.013052
Amazon Echo	0.012345
Belkin Netcam	0.12345
Logitech Hurmony Hub	0.011895
Bose SoundTouch 10	0.01888
iPhone	0.011069
August doorbell cam	0.008807
Belkin WeMo Link	0.008807
Belkin WeMo Motion Sensor	0.008807
Belkin WeMo Switch	0.008807
Canary	0.008807
Caseta Wireless Hub	0.008807
Chamberlain myQ garage opener	0.008807
Harmon Kardon Invoke	0.008807
Insteon Hub	0.008807
Koogeek Lightbulb	0.008807
LIFT Virtual Bulb	0.008807
Logitech Logi Circle	0.008807
Nest Camera	0.008807
Nest Cam IQ	0.008807
Nest Quard	0.008807
Netgear Arlo Camera	0.008807
nVidia Shield	0.008807

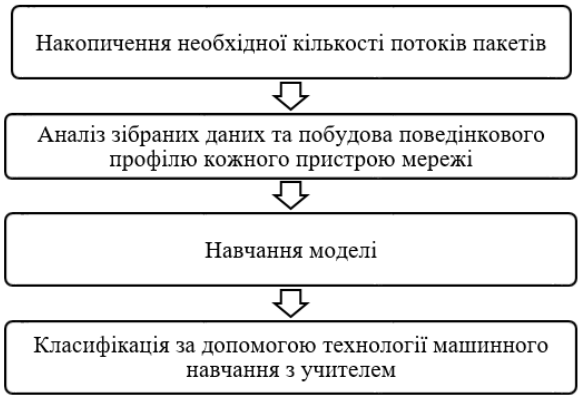

Рис. 2. Алгоритм методу

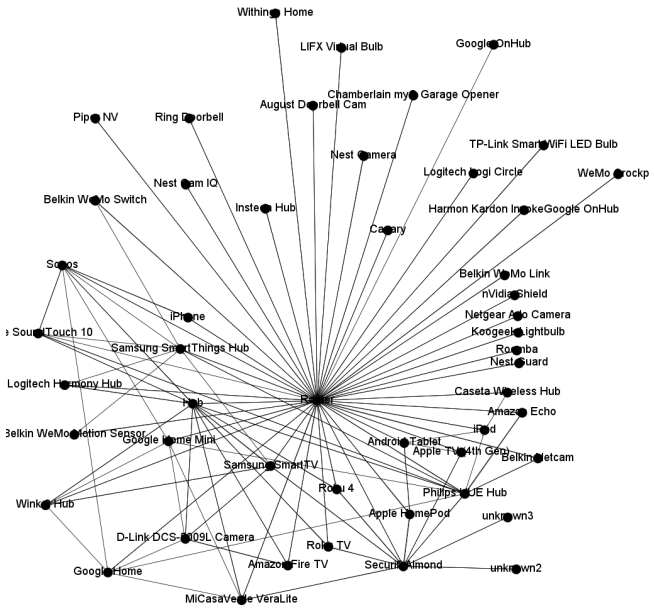

Рис. 3. Зв'язки між пристроями в локальній мережі

6) Кількість TCP з'єднань з IP адресами, що належать одному домену.
7) Кількість всіх TCP з'єднань.
8) Кількість TCP з'єднань в локальній мережі.
9) Сила зв'язності пристроїв локальної мережі.

Для ідентифікації пристроїв було обрано технологію випадкового лісу – ансамблевий метод машинного навчання для класифікації, який оперує за допомогою побудови чисельних дерев прийняття рішень під час тестування моделі і продукує моду для класів. Причиною вибору випадкового лісу є його висока стійкість до перенаглагодження порівняно з іншими класифікаторами дерева рішень.

3. Практичне застосування методу

Для проведення необхідних експериментів було обрано два незалежні набори даних (датасети) [8], [9], кожен з яких містив у собі як вузькоцільові так і багатоцільові пристрої Інтернету речей. Для тестування було обрано датасет [8], що включав у собе 20 багатоцільових та 34 вузькоцільових пристроїв. Структуру мережі та взаємозв'язків пристроїв для обраного датасету зображено на рис. 3. Для отримання більш точних поведінкових профілів для кожного пристрою було обрано період спостереження, що склав 4 дні.

Для валідації методу було обрано датасет [9].

Варто зазначити, під час спостереження було виявлено оптимальну межу в 21000 перших зібраних пакетів для кожного пристрою, за допомогою яких вдалося отримати 94% точності. В загальному випадку, було виявлено 9 багатоцільових пристроїв та 22 вузькоцільових пристроїв зі спростережуваної мережі.

Результат класифікації, що наведено в таблиці 2 демонструє, що технологія випадкового лісу досягає стабільно високої точності, що сягає 94% в різних мережах. Хибними виявилися два рішення, що від-

Табл. 2. Результат класифікації з встановленою межею в 21000 пакетів

Передбачено / Фактично	Вузькоцільовий пристрій	Багатоцільовий пристрій
Вузькоцільовий пристрій	22	0
Багатоцільовий пристрій	2	7

Табл. 3. Результат класифікації з встановленою межею в 560000 пакетів

Передбачено / Фактично	Вузькоцільовий пристрій	Багатоцільовий пристрій
Вузькоцільовий пристрій	22	0
Багатоцільовий пристрій	1	8

несли акустичні системи з підвищеною вбудованою функціональністю до вузькоцільових пристроїв, в силу пасивної (неактивної) поведінки в мережі. Однак, при підвищенні межі кількості пакетів для класифікації до 560000, отримуємо вього одну помилку (табл. 3), при цьому точність зростала до 97%. Також важливо помітити, що в обох результатах відсутні помилки, пов'язані з класифікуванням вузькоцільових пристроїв як багатоцільових. Це означає, що даний підхід здатен виявити за допомогою поведінкових профілів потоку мережевих пакетів кожного пристрою вузькоцільові та багатоцільові пристрої в різних мережах.

Висновки

Результатом досліждення було представлено один з можливих способів удосконалення методу класифікації вузькоцільових та багатоцільових пристроїв Інтренету речей, застосовуючи додаткові припущення. Як наслідок, різниця в точності класифікації пристроїв між набором даних для тестування і валідації зменшена до 3% в порівнянні з 10% в моделі, запропонованій раніше. Валідація на різних наборах даних показує, що модель більш стійка до різних мереж. Отриманий результат може бути використаний в подальшому розмежуванні пристроїв для більш гнучкого формування політик безпеки, створення віртуальних мереж та досягнення вищої якості обслуговування і унеможливлюванні взаємних перешкоджень в роботі.

Ця стаття дає поштовх майбутнім дослідженням у сфері безпеки локальних мереж від несанкціонованого впливу на пристрої, а також в сфері забезпечення необхідної продуктивності та якості обслуговування в середовищі Інтернету речей.

Перелік використаних джерел

1. Dennis Knake IoT numbers vary drastically: devices and spending in 2020 — 2016. — Access mode: `https://www.wespeakiot.com/iot-numbers-devices-spendings-2020`.
2. G. Mezzofiore A university was attacked by its lightbulbs, vending machines and lamp posts — 2017. — Access mode: `https://mashable.com/2017/02/13/internet-of-things-university-network-/#9RqajU3A5Oqu`.
3. Cisco Cisco 2017 Midyear Cybersecurity Report — 2017. — Access mode: `https://www.cisco.com/c/dam/global/es_mx/solutions/security/pdf/cisco-2017-midyear-cybersecurity-report.pdf`.
4. Thien Duc Nguyen, S. Marchal, M. Miettinen, H. Fereidooni, N. Asokan, Ahmad-Reza Sadeghi DIOT: A Federated Self-learning Anomaly Detection System for IoT. — 2019. — Access mode: `https://export.arxiv.org/pdf/1804.07474`.
5. H. Guo, J. Heidemann IP-Based IoT Device Detection — 2018. — Access mode: `https://www.isi.edu/~johnh/PAPERS/Guo18b.pdf`.
6. A. Sivanathan, D. Sherratt, H. Gharakheili, A. Radford, C. W. Vishwanath, V. Sivaraman Characterizing and Classifying IoT Traffic in Smart Cities and Campuses — 2018. — Access mode: `http://www2.ee.unsw.edu.au/~vijay/pubs/conf/17infocom.pdf`.
7. S.Brin, L. Paghan The anatomy of a large-scale hypertextual Web search engine — 1998. — Access mode: `http://snap.stanford.edu/class/cs224w-readings/Brin98Anatomy.pdf`.
8. A. Sivanathan, H. Habibi Gharakheili, F. Loi, A. Radford, C. Wijenayake, A. Vishwanath and V. Sivaraman Classifying IoT Devices in Smart Environments Using Network Traffic Characteristics — 2018. — Retrieved from: `https://iotanalytics.unsw.edu.au/iottraces`.
9. O. Alrawi, C. Lever, M. Antonakakis, F. Monrose SoK: Security Evaluation of Home-Based IoT Deployments — 2018. — Retrieved from: `https://yourthings.info/data/`.